Transactions of the American Fisheries Society 124:911–928, 1995

Properties of Catch Rates Used in Analysis of Angler Surveys

C. M. JONES

Applied Marine Research Laboratory
Old Dominion University, Norfolk, Virginia 23529-0456, USA

D. S. ROBSON

150 MacClaren Street, Penthouse 6, Ottawa, Ontario K2P 0L2, Canada

H. D. LAKKIS[1]

Applied Marine Research Laboratory and Department of Mathematics and Statistics
Old Dominion University, Norfolk, Virginia 23529-0456, USA

J. KRESSEL[2]

Applied Marine Research Laboratory

Abstract.—On-site angler surveys commonly yield data on fish catch, fishing effort, and their variances for a sample of anglers. The ratio of catch to effort, or catch rate, often is multiplied by an independent estimate of total effort to calculate total catch (with its confidence interval) throughout the fishery. The most frequently used measures of catch rate are the ratio-of-means estimator (mean catch per angler divided by mean effort per angler) and the mean-of-ratios estimator (mean angler's catch rate). Bias and misleading confidence intervals are associated with use of ratio estimators, and the best catch rate measure for estimating total catch has been uncertain. We used statistical theory and simulation modeling to demonstrate that the most appropriate estimator (least bias, truest confidence interval) depends on the method of sampling. The ratio-of-means catch rate is better when anglers are sampled with equal probability at the completion of their trips (as in access point surveys). The mean-of-ratios estimator should be used when anglers are sampled, while still fishing, with probabilities proportional to trip length (as in roving surveys). When either estimator is used, confidence intervals around total catch are strongly influenced by skewness in the distribution of catch rates among anglers. At least 100 anglers or angler parties must be interviewed to attain targeted two-tailed α-levels; even then, the two statistical rejection areas (ideally, $\alpha/2$) differ between tails of the distribution. Our results clarify the appropriate use of catch rate estimators with survey data, and they show the desirability of larger sample sizes than are customarily used in angler surveys.

The estimation of total catch is amongst the most important objectives in many angler surveys. Yet a direct estimate of total catch is not possible in recreational fisheries when the access to the waterbody is from private property or where sites are too numerous to survey (Malvestuto 1983; Hayne 1991; Robson 1991; Pollock et al. 1994). In such circumstances, total catch can only be estimated from mean catch rate, which is a ratio estimate, and an independent measure of total effort—commonly expressed as $\hat{Y} = \hat{R} \times \hat{X}$, where Y is total catch, R is catch rate, and X is total effort.

The use of ratio estimates in recreational fisheries, however necessary, presents several difficulties (Thompson 1992; Pollock et al. 1994). Two problems are of particular interest in estimating catch. First, there are several ways to calculate catch rate, and the literature reveals confusion about the correct method of calculation. Second, bias and incorrect confidence interval coverage are inherent in the use of ratios. Both of these problems are exacerbated by the skewness often seen in frequency plots of catch and catch rates.

The use of mean catch rate in estimating total catch is actually more complex than first appears. As Brown (1971) and Crone and Malvestuto (1991) pointed out, there are several ways to calculate mean catch rate. Two estimators used are the mean of the ratios (which we call the "per-angler" estimator) and the ratio of the means (which we call the "per-day" estimator). The per-angler estimator is calculated as the average angler's catch rate, the per-day estimator by dividing

---

[1] Present address: Center for Pediatric Research, 855 West Brambleton Avenue, Norfolk, Virginia 23510-1001, USA.

[2] Present address: National Aeronautics and Space Administration, Goddard Space Flight Center, Code 664, T-2, Room 78, Greenbelt, Maryland 20771, USA.

Methods

We first provide formulae for catch rate parameters that are estimated, the commonly used ratio estimators, and regression estimators (alternative forms of the ratio estimators that show potential bias), and we give exact variance formulae for these estimators. Second, we use these definitions to show how confidence intervals and their associated t-values are calculated for the simulation model. Finally, we present the data and decision rules used in the model.

Notations and Parameters

Let $i = 1, 2, \ldots, N$, x_i = trip length of the ith angler or party in hours (fishing effort), y_i = catch by the ith angler or party, n = number of anglers or parties interviewed, and N = number of anglers or parties in the fishery on a given day. The following sample statistics are used in derivations:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \text{sample mean of angler or party effort;}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \text{sample mean of angler or party catch;}$$

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} = \text{sample variance of angler or party effort;}$$

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1} = \text{sample variance of angler or party catch.}$$

We define the ordinary population parameters that will be used in derivations of the sampling property of estimators of catch rate as:

$$X = \sum_{i=1}^{N} x_i = \text{total effort of all anglers or parties in the fishery on the day in question;}$$

$$Y = \sum_{i=1}^{N} y_i = \text{total catch of all anglers or parties in the fishery;}$$

$$\mu_x = \frac{X}{N} = \text{mean of angler or party effort;}$$

$$\mu_y = \frac{Y}{N} = \text{mean of angler or party catch;}$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu_x)^2 = \text{variance of angler or party effort;}$$

$$C_x \equiv \frac{\sigma_x}{\mu_x} = \text{coefficient of variation of } x\text{;}$$

$$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N} (y_i - \mu_y)^2 = \text{variance of angler or party catch;}$$

$$C_y \equiv \frac{\sigma_y}{\mu_y} = \text{coefficient of variation of } y\text{;}$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \equiv \rho_{xy}\sigma_x\sigma_y = \text{covariance of } y \text{ and } x\,;$$

$$\beta = \frac{\sigma_{xy}}{\sigma_x^2} = \rho_{xy}\frac{\sigma_y}{\sigma_x} = \text{slope of the ordinary least-squares (OLS) regression of catch on effort;}$$

$$\delta = \mu_y - \beta\mu_x = \text{intercept of the OLS regression of catch on effort.}$$

When the probability of encountering an angler is proportional to trip length and is therefore weighted by effort, the population parameters that appear in derivations of sampling estimators of catch rate are defined as:

$$w_i = \frac{x_i}{X} = \text{sampling probability of the } i\text{th angler or party;}$$

$$\mu_{x,w} = \sum_{i=1}^{N} w_i\,x_i = \text{weighted mean of angler or party effort;}$$

$$\mu_{y,w} = \sum_{i=1}^{N} w_i y_i = \text{weighted mean of angler or party catch;}$$

$$\sigma_{x,w}^2 = \sum_{i=1}^{N} w_i(x_i - \mu_{x,w})^2 = \text{weighted variance of angler or party effort;}$$

$$C_{x,w} = \frac{\sigma_{x,w}}{\mu_{x,w}} = \text{coefficient of weighted variation of effort;}$$

$$\sigma_{y,w}^2 = \sum_{i=1}^{N} w_i(y_i - \mu_{y,w})^2 = \text{weighted variance of angler or party catch;}$$

$$\sigma_{xy,w} = \sum_{i=1}^{N} w_i(y_i - \mu_{y,w})(x_i - \mu_{x,w}) = \text{weighted covariance of angler or party catch and effort;}$$

$$\beta_w = \frac{\sigma_{xy,w}}{\sigma_{x,w}^2} = \rho_{xy,w}\frac{\sigma_{y,w}}{\sigma_{x,w}} = \text{slope of the weighted least-squares regression of catch on effort;}$$

$$\delta_w = \mu_{y,w} - \beta_w\mu_{x,w} = \text{intercept of the weighted regression of catch on effort.}$$

Catch Rate Parameters and Regression Formulation

The two most common catch rates of interest in angler surveys are

$$R_1 = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} = \text{per-day catch rate (the means are implicit, the sample sizes having canceled out) and}$$

$$R_2 = \frac{\sum_{i=1}^{N} \frac{y_i}{x_i}}{N} = \text{per-angler or per-party catch rate.}$$

We include the formulation for a third parameter, R_3. This parameter is not a useful catch rate parameter, but it is estimated inadvertently and unknowingly in many creel surveys, as we show in the results section. It is defined as

$$R_3 = \frac{\mu_{y,w}}{\mu_{x,w}} = \text{weighted per-day catch rate.}$$

Fitted regressions are used to produce estimates of the intercept, δ, which relates to the sign and magnitude of bias. The ordinary least-squares regression line of catch on effort, $y_i = \beta x_i + \delta$, relates to the per-day catch rate when sampling is with equal probability. The weighted (by w_i) least-squares regression line $y_i = \beta_w x_i + \delta_w$ relates to the per-day catch rate when sampling is with probability proportional to trip length.

Catch Rate Estimators

The intuitive formulas for the estimators of R_1 and R_2 are

$$\hat{R}_1 = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}, \text{ the per-day estimator, and} \tag{1}$$

$$\hat{R}_2 = \frac{\sum_{i=1}^{n} \frac{y_i}{x_i}}{n}, \text{ the per-angler estimator.} \tag{2}$$

The approximate expected values for these estimators for sampling with equal probability and with probability proportional to trip length were derived with Taylor series expansions. The derivation of the per-angler estimator ($\hat{R}_2$) for sampling proportional to trip length is presented in Appendix 1. Because this derivation shows that $\hat{R}_2$ actually estimates R_1 under these conditions, we will now call this estimator $\tilde{R}_1$ to indicate the true parameter that it estimates. Additional properties of R_1 are shown in Appendix 2.

Variance Formulas

The approximate variance of the per-day estimator (equation 1) for equal probability sampling is given (Cochran 1977:253) as

$$V(\hat{R}_1) \cong \frac{1}{n(\bar{X})^2}\left[\frac{\sum_{i=1}^{N} (y_i - R_1 x_i)^2}{N}\right]. \tag{3}$$

The exact variance of the per-angler estimator for sampling proportional to trip length (equation 2) is given (Cochran 1977:253) as

$$V(\tilde{R}_1) = \frac{\sum_{i=1}^{N} x_i\left(\frac{y_i}{x_i} - R_1\right)^2}{n\sum_{i=1}^{N} x_i}. \tag{4}$$

The exact variance for the per-angler estimator for equal probability sampling (equation 2) is given as

$$V(\hat{R}_2) = \frac{\sum_{i=1}^{N}\left(\frac{y_i}{x_i} - R_2\right)^2}{nN}. \tag{5}$$

Confidence Intervals

To estimate the confidence interval coverage, we simulated the sampling distribution of the actual t-values for each estimator, $\hat{R}_1$, $\tilde{R}_1$, and $\hat{R}_2$. We compared the simulated t-values with the tabulated (or targeted) values from the Student's t-distribution. For the per-day estimator $\hat{R}_1$ and equal probability sampling, the simulated t is calculated from Fieller's theorem (Cochran 1977:156) as

$$t_1 = \frac{\bar{y} - R_1\bar{x}}{\mathrm{SE}(\bar{y} - R_1\bar{x})}, \quad \text{where } \mathrm{SE}(\bar{y} - R_1\bar{x}) = \sqrt{\frac{s_y^2 + R_1^2 s_x^2 - 2R_1 s_{xy}}{n}}, \quad \text{and where}$$

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1}.$$

The confidence interval for R_1, based on $\hat{R}_1$, is

$$\frac{\overline{xy} - \frac{t_{\alpha/2}^2 s_{xy}}{n} \pm \sqrt{\left(\overline{xy} - t_{\alpha/2}^2 \frac{s_{xy}}{n}\right)^2 - \left(\bar{x}^2 - t_{\alpha/2}^2 \frac{s_x^2}{n}\right)\left(\bar{y}^2 - t_{\alpha/2}^2 \frac{s_y^2}{n}\right)}}{\bar{x}^2 - t_{\alpha/2}^2 \frac{s_x^2}{n}}.$$

The above confidence interval covers R_1 only if t_1 satisfies $-t_{\alpha/2} < t_1 < t_{\alpha/2}$.

For the per-angler estimator $\tilde{R}_1$ and sampling proportional to trip length, the simulated t is calculated as

$$t_2 = \frac{\tilde{R}_1 - R_1}{\mathrm{SE}(\tilde{R}_1)}, \text{ where } \mathrm{SE}(\tilde{R}_1) = \sqrt{\frac{\sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \tilde{R}_1\right)^2}{n(n - 1)}}.$$

The confidence interval for R_1, based on $\tilde{R}_1$, is $\tilde{R}_1 \pm t_{\alpha/2}\mathrm{SE}(\tilde{R}_1)$, and the estimated probability of coverage of R_1 based on $\tilde{R}_1$ is the empirical proportion of sample t_2-values satisfying

$-t_{\alpha/2} < t_2 < t_{\alpha/2}$.

For the per-angler estimator $\hat{R}_2$, the simulated t is calculated as

$$t_3 = \frac{\hat{R}_2 - R_2}{\mathrm{SE}(\hat{R}_2)}, \quad \text{where } \mathrm{SE}(\hat{R}_2) = \sqrt{\frac{s_{y/x}^2}{n}}, \quad \text{and where}$$

$$s_{y/x}^2 = \frac{\sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \frac{\sum_{i=1}^{n} \left(\frac{y_i}{x_i}\right)}{n}\right)^2}{n - 1}.$$

The confidence interval for R_2 is $\hat{R}_2 \pm t_{\alpha/2}\mathrm{SE}(\hat{R}_2)$, and the estimated probability of coverage of R_2 is the empirical proportion of $\hat{R}_2$ values satisfying $\hat{R}_2 - t_{\alpha/2}\mathrm{SE}(\hat{R}_2) < R_2 < \hat{R}_2 + t_{\alpha/2}\mathrm{SE}(\hat{R}_2)$, which is the same as the proportion of sample t_3-values satisfying $-t_{\alpha/2} < t_3 < t_{\alpha/2}$.

We see in the computation of t_1, t_2, and t_3 that either the numerator or denominator can be responsible for any asymmetry seen in t-distributions generated through simulation. Hence, it is instructive to rewrite the numerator of the t-values as a function of the approximately normal random variate and the denominator as a function of the approximately chi-square (χ^2) variate. We can then compare these components from our simulated t-distributions with their associated theoretical curves. Calculations used to transform numerators and denominators of the simulated t-variates for comparison with theoretical Student's t-variates are given in Appendix 3.

Simulation Model

The simulation model was used to investigate the approach of realized α-levels to targeted α-levels and to determine asymmetry in upper and lower confidence bands. We obtained data from a 1989 survey of anglers in the Virginia fishery for summer flounder

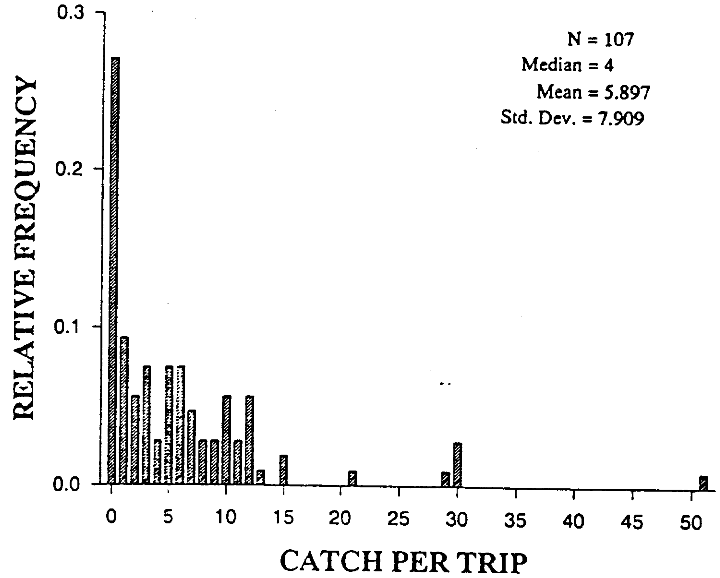


FIGURE 1.—Empirical distribution of the number of fish caught per trip by angling parties fishing for summer flounder in Virginia, 1989. Sample size (N) was 107 parties; std. dev. is standard deviation.

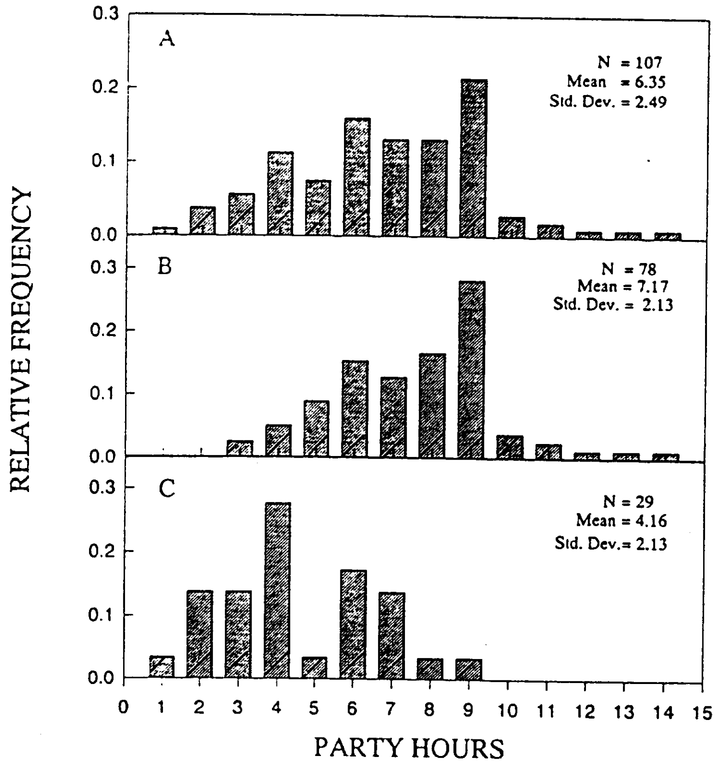


FIGURE 2.—Distributions of angling effort among 107 parties that had fished for summer flounder (Figure 1): (A) all parties; (B) parties successful in catching fish; (C) parties unsuccessful in catching fish. Successful anglers fished longer than unsuccessful anglers; std. dev. is standard deviation.

Paralichthys dentatus (Jones et al. 1990). The 107 parties interviewed had completed their trips and had fished specifically for summer flounder. The distributions of catch and effort are plotted in Figures 1 and 2. In this fishery, 10% of the anglers caught 71% of the fish, which is a typical pattern. For these data, the per-day catch rate R_1 was 0.928 fish/h and the per-angler catch rate R_2 was 0.802 fish/h.

For each simulation, we randomly drew a predetermined number of angling parties from the empirical population and logged the catch and effort of each. For the two estimators $\hat{R}_1$ and $\hat{R}_2$, the parties in each sample were selected with replacement, which effectively made the population infinite. For $\tilde{R}_1$, the parties in each sample were selected with replacement and with probability proportional to trip length.

Four levels of sampling intensity were simulated by selecting 10, 30, 60, and 100 parties. This range of coverage allowed us to estimate the necessary sample size at which targeted and realized α-levels converged. These party levels bracketed the anticipated sampling level (30 parties) at which normality is usually reached even with skewed distributions (Thompson 1992). For each of these four fishing intensity levels, 10,000 iterations were run, producing 10,000 values of t and mean catch rate.

Results

Correct Choice of Catch Rate

The two sampling schemes (with probability equal or proportional to trip length) affect the derivations of expected value and the population parameter actually estimated (Table 1). When we use the per-day catch rate estimator $\hat{R}_1$ and sample with equal probability, we obtain a biased estimate of R_1. This bias is on the order of $1/n$ and its sign depends on the sign of the intercept (δ) of the ordinary least-squares regression line. When we use $\hat{R}_1$ and sample proportional to trip length, we obtain a biased estimate of a new parameter that we call R_3 (Table 1). This bias also is on the order of $1/n$ and its sign depends on the sign of the intercept (δ_w) of the weighted least-squares regression line. This quantity is ostensibly a catch rate estimator, but does not estimate either the per-day (R_1) or the per-angler (R_2) population catch rates. Indeed, it is a quantity that is not of use to fishery scientists. When we use the per-angler catch rate estimator $\hat{R}_2$ and sample with equal probability, we obtain an unbiased estimate of R_2, as expected. However, when we use the same estimator but sample with probability proportional to trip length, we obtain an unbiased estimate of the per-day catch rate R_1.

Relationship between Measures of Catch Rate

The true ratios (R_1, R_2, R_3) model different relationships in these data (Figure 3). The plots of $y = R_1x$, $y = R_2x$, and $y = R_3x$ show that within any day, $R_3 > R_1 > R_2$. In Appendix 2 we show the mathematical relationship between R_1 and R_3. Even though R_3 can be derived, it is not a useful quantity in evaluating recreational fisheries, as already noted (Table 1). The results show that for these data, the per-day ratio is intrinsically larger than the per-angler ratio.

Linear regressions showed that the estimator $\hat{R}_1$ has negative bias whether sampling probabilities were equal or proportional to effort (Figure 3). Both the ordinary least-squares regression (OLS: $y_i = \beta x_i + \delta$) used for equal sampling probabilities and the weighted least-squares regression (WLS: $y_i = \beta_w x_i + \delta_w$) used for unequal probabilities had negative intercepts, indicating negative bias in the $\hat{R}_1$ estimators. These biases arose because zero catches of summer flounder were common (27%), a likely sportfishing outcome. When zero catches occur at lower levels of effort, $\hat{R}_1$ will underestimate R_1 for equal probability sampling and R_3 for sampling probabilities proportional to effort. In contrast, regressions of the form $y_i = Rx_i$ pass through the origin.

Although we have dealt here with within-day estimates and do not evaluate seasonal measures of catch rate, a negative correlation may exist between total annual sport catch and total annual sport effort, as is commonly the case in commercial fisheries. In this case the relationships shown in Figure 3 would also hold for the seasonal ratio estimators.

Confidence Coverage

Per-day (ratio of means) estimator $\hat{R}_1$ and sampling with equal probability.—For $\hat{R}_1$ and equal probability sampling, realized α-levels approached targeted α-levels more slowly than they did with other estimators (Table 2). Table 2 shows results for three levels of significance but, for simplicity,

TABLE 1.—Relationships between the statistical estimators of catch rate, the sampling scheme, and the population parameters that are estimated. The cells of this table contain the expected values of the estimators $\hat{R}_1$ and $\hat{R}_2$ under two sampling schemes. When sampling is done with equal probability, the per-day estimator $\hat{R}_1$ provides a biased estimate of R_1. When sampling is done with probability proportional to angler trip length, $\hat{R}_1$ provides a biased but consistent estimate of R_3, the mean weighted catch divided by the mean weighted effort. However, R_3 is not of interest to fishery scientists because it does not equal either R_1 or R_2. When sampling is done with equal probability, the per-angler estimator $\hat{R}_2$ provides an unbiased estimate of R_2 for the population. When sampling is proportional to trip length, $\hat{R}_2$ provides an unbiased estimate of the per-day catch rate R_1 for the population.

Catch rate estimator	Sampling with equal probability and with replacement	Sampling with probability proportional to trip length and with replacement
Per day, $\hat{R}_1 = \dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$	Estimator $\hat{R}_1$ is an estimate of $R_1 + \dfrac{\delta}{n\mu_x}\left(\dfrac{\sigma_x}{\mu_x}\right)^2$	Estimator $\hat{R}_1$ is an estimate of $R_3 + \dfrac{\delta_w}{n\mu_{x,w}}\left(\dfrac{\sigma_{x,w}}{\mu_{x,w}}\right)^2$
Per angler, $\hat{R}_2 = \dfrac{\sum_{i=1}^{n} \frac{y_i}{x_i}}{n}$	Estimator $\hat{R}_2$ is an estimate of R_2	Estimator $\hat{R}_2$ (now called $\tilde{R}_1$) is an estimate of R_1

we limit this discussion to the case of targeted $\alpha = 0.05$, the traditional level for 95% confidence coverage. The realized α-level obtained from $\hat{R}_1$ simulations was higher than the targeted value of 0.05 for 10 interviews ($\alpha = 0.135$), for 30 interviews ($\alpha = 0.106$), and for 60 interviews ($\alpha = 0.086$). In these cases, the true confidence coverages would only be 86.5%, 89.4%, and 91.4%. At 100 interviews, the realized α (0.067) was still not equal to the targeted value and yielded a confidence coverage of 93.3% rather than the expected 95%.

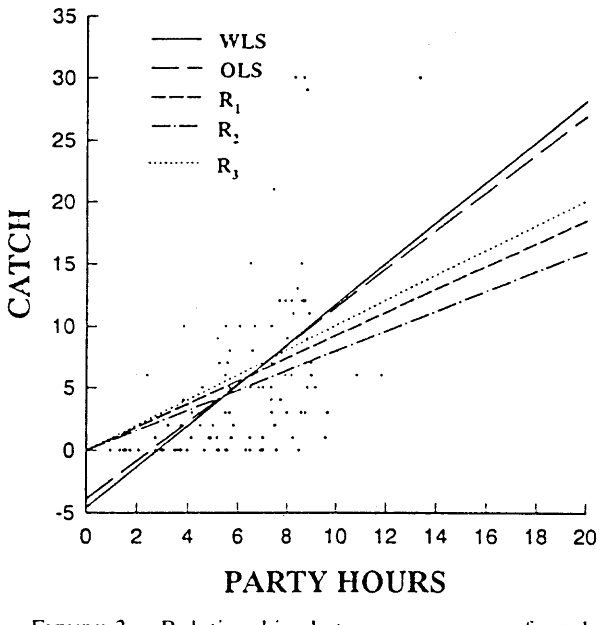


FIGURE 3.—Relationships between measures of catch rate R_1 (total catch/total effort: the per-day catch rate), R_2 (the per-angler catch rate), and R_3 (a frequently used, but inappropriate measure). The values of these measures for the summer flounder data (dots) are $R_1 = 0.928$, $R_2 = 0.802$, and $R_3 = 1.0095$. Also illustrated are the ordinary least-squares (OLS: $y_i = 1.5465x_i - 3.93$) and weighted least-squares (WLS: $y_i = 1.6406x_i - 4.618$) regression lines (y is catch; x is hours of effort). The intercepts of these regressions show the magnitude of bias in the per-day estimator $\hat{R}_1$ when sampling is done with equal probability (OLS) and when sampling is done proportional to trip length (WLS).

The t-values (Table 2; Figures 4–7) for this estimator were not symmetrical about a mean of 0 as would be the case with Student's t-distribution. The rejection area was concentrated in the left tail, which contained 87–99% of the total. We saw t-values smaller than −8 with a sample size of 10 interviews. The left tail remained too large even with 100 interviews, containing 5.8% of the total area compared to the expected 2.5%.

Per-angler estimator $\tilde{R}_1$ and sampling proportional to trip length.—For $\tilde{R}_1$ and sampling probability proportional to trip length, α-levels approached normality faster than they did with estimator $\hat{R}_1$. The results of the simulation for estimator $\tilde{R}_1$ of the per-day catch rate are shown in Table 2 and Figures 4–7. Realized α-levels obtained from simulations were again higher than the targeted value of 0.05 for 10 ($\alpha = 0.101$), 30 ($\alpha = 0.079$), 60 ($\alpha = 0.065$), and 100 interviews ($\alpha = 0.062$). The corresponding true confidence coverages would only be 89.9%, 92.1%, 93.5%, and 93.8%.

As with the per-day estimator $\hat{R}_1$, the t-values of $\tilde{R}_1$ were not symmetrical (Table 2; Figures 4–7). The rejection area was again concentrated in the left tail, which contained 80–96% of the total. We saw t-values smaller than −7 with a sample

TABLE 2.—Comparisons of targeted and realized α probabilities in relation to the simulated number (N) of anglers interviewed. The targeted values are taken from Student's t-distribution with $N - 1$ df. The realized values (expressed as percentages) are taken from 10,000 simulated t-values as the proportion of $t < -t_{\alpha/2}$ added to the proportion of $t > t_{\alpha/2}$, where $t_{\alpha/2}$ is the cut point from the Student t-distribution at the targeted $\alpha/2$. The SEs of $\hat{R}_1$, $\tilde{R}_1$, and $\hat{R}_2$ were calculated by taking the square roots of text equations (3), (4), and (5), respectively. Confidence intervals are denoted CIs.

		Targeted α = 1%; % of CIs:		Targeted α = 5%; % of CIs:		Targeted α = 10%; % of CIs:	
N	Estimator SE	Below true R	Above true R	Below true R	Above true R	Below true R	Above true R
Per-day estimator $\hat{R}_1$: sampling with equal probability and $R = R_1$ in the confidence bounds							
10	0.277	5.9	0.0	13.3	0.2	18.8	0.7
30	0.160	5.5	0.0	10.3	0.3	14.3	1.3
60	0.113	3.7	0.1	8.0	0.6	11.1	1.7
100	0.088	2.5	0.1	5.8	0.9	9.0	2.4
Per-angler estimator $\tilde{R}_1$: sampling proportional to trip length and $R = R_1$ in the confidence bounds							
10	0.318	4.2	0.1	9.7	0.4	14.2	1.2
30	0.184	2.9	0.1	7.3	0.6	11.0	2.0
60	0.130	2.0	0.1	5.6	0.9	8.9	2.5
100	0.101	1.9	0.1	5.0	1.2	8.0	3.1
Per-angler estimator $\hat{R}_2$: sampling with equal probability and $R = R_2$ in the confidence bounds							
10	0.304	3.9	0.0	9.8	0.0	13.6	0.4
30	0.176	3.3	0.0	6.9	0.4	10.8	1.9
60	0.124	2.1	0.0	5.2	1.1	8.8	2.5
100	0.096	1.8	0.0	5.2	0.7	0.095	0.03

size of 10 interviews. The left tail remained too large even with 100 interviews, containing 5.0% of the area of the distribution instead of the expected 2.5%.

Per-angler (mean of ratios) estimator $\hat{R}_2$ and sampling with equal probability.—For $\hat{R}_2$, realized α-levels again did not approach targeted levels as expected when the sample size became as large as 30 interviews (Table 2). Relative to a target α of 0.05, realized α-levels obtained from the simulations were high for 10 (α = 0.098), 30 (α = 0.073), and 60 interviews (α = 0.063). In these cases, the true confidence coverages would only be 90.0%, 92.7%, and 93.7% rather than the 95% expected. At 100 interviews, the targeted (α = 0.05) and realized (α = 0.059) values were close.

The t-values again were not symmetrical about a mean of 0 (Table 2; Figures 8–11); 88–100% of the rejection area was concentrated in the left tail of the empirical t-distribution. With a sample size of 10 interviews, we saw t-values smaller than −10 in the left tail (Figure 8). This skewness improved as sample size increased to 100, but even at this level the left tail contained 5.2% of the total curve area rather than the expected 2.5%.

Source of Asymmetry in Empirical t-Distributions

Comparisons of the numerators and denominators for the three estimators showed that the problem of skewness was caused by the denominators (Figures 4–11). The distribution of the numerator for all three estimators was almost identical to that of the theoretical z-value by the time the sample size reached 60 (Figures 6, 10). In contrast, the simulated cumulative frequency distribution of the denominator was erratic for small sample sizes, and although it became smoother as sample size increased, it was still quite far from the theoretical χ^2 distribution at our largest sample size. The denominator of the per-angler estimator $\hat{R}_2$ (equal sampling probability) was slightly closer to the theoretical χ^2 than the estimators of the daily catch rate, $\hat{R}_1$ and $\tilde{R}_1$. This explains why the realized α-values of $\hat{R}_2$ approached the targeted values faster than they did with the other two estimators, for all significance levels. Likewise realized α-values approached targeted values faster with $\tilde{R}_1$ than with $\hat{R}_1$.

Discussion

Superficially, calculation of catch rates and their confidence intervals in recreational fisheries appears straightforward. However, the choice of calculation method is actually dictated by the objective of the study and the sampling regime. Creel surveys are designed to sample anglers either with equal probability (access point method) or with probability proportional to trip length (roving method). The objective of creel surveys is most

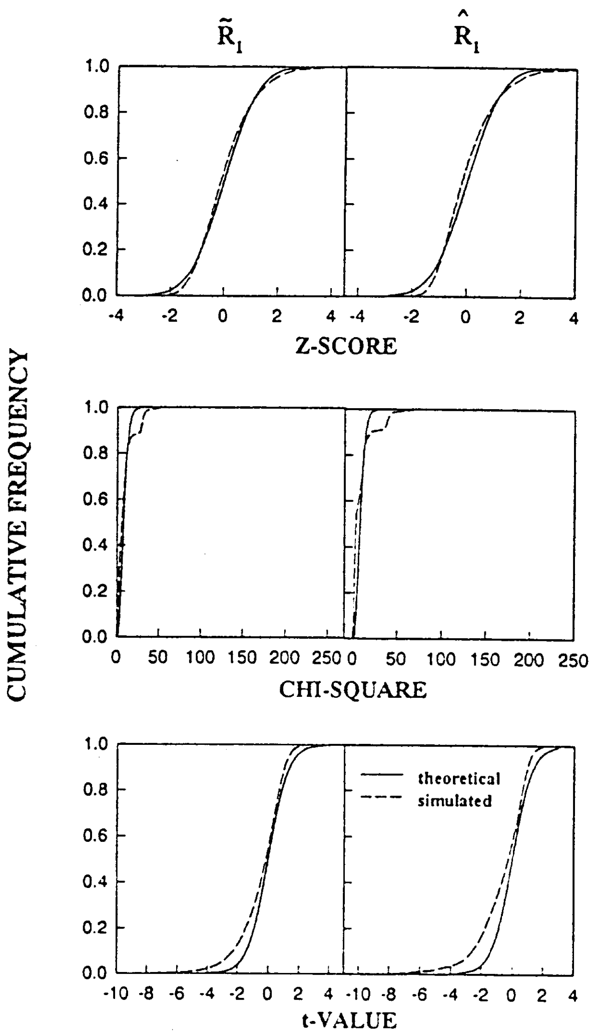


FIGURE 4.—Properties of the distributions of the per-day catch rate estimators $\hat{R}_1$ and $\tilde{R}_1$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerators (z), denominators (χ^2), and t-values for samples of 10 interviews.

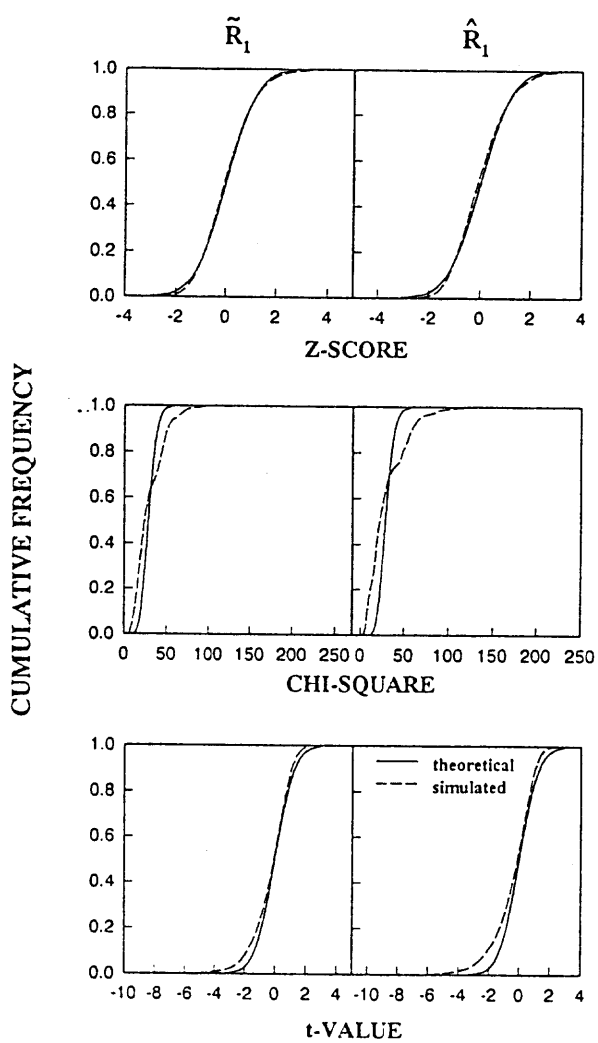


FIGURE 5.—Properties of the distributions of the per-day catch rate estimators $\hat{R}_1$ and $\tilde{R}_1$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerators (z), denominators (χ^2), and t-values for samples of 30 interviews.

often to estimate catch and effort, but occasionally surveys are used to indicate the quality of anglers' fishing experiences.

Total catch can be calculated by either of two estimators. When anglers are sampled with equal probability for a day's entire sample, $\hat{R}_1$ is the sample total catch divided by the sample total effort, and it estimates the corresponding per-day catch rate (R_1) for the entire fishery. It can easily be combined with an independent estimate of effort to provide an unbiased estimate of total catch. When anglers are sampled in proportion to their trip length, $\hat{R}_1$ becomes an estimator of R_3, the weighted per-day catch rate; R_3 is typically larger than R_1 and leads to an overestimate of catch when multiplied by an independent estimate of effort. When anglers are sampled in proportion to their trip length, the per-angler estimator $\hat{R}_2$, now called $\tilde{R}_1$, becomes unbiased for the fishery's per-day catch rate (R_1), and it can be multiplied by an independent estimate of effort to provide a consistent estimate of catch (unbiased for large sample sizes). Finally, when anglers are sampled with equal probability, the simple arithmetic mean ($\hat{R}_2$) of catch rates in the sample is an unbiased estimator of the per-angler catch rate R_2 in the population; however, although $\hat{R}_2$ has the well-known statistical properties of sample means, it does not provide a statistically "consistent" estimate of to-

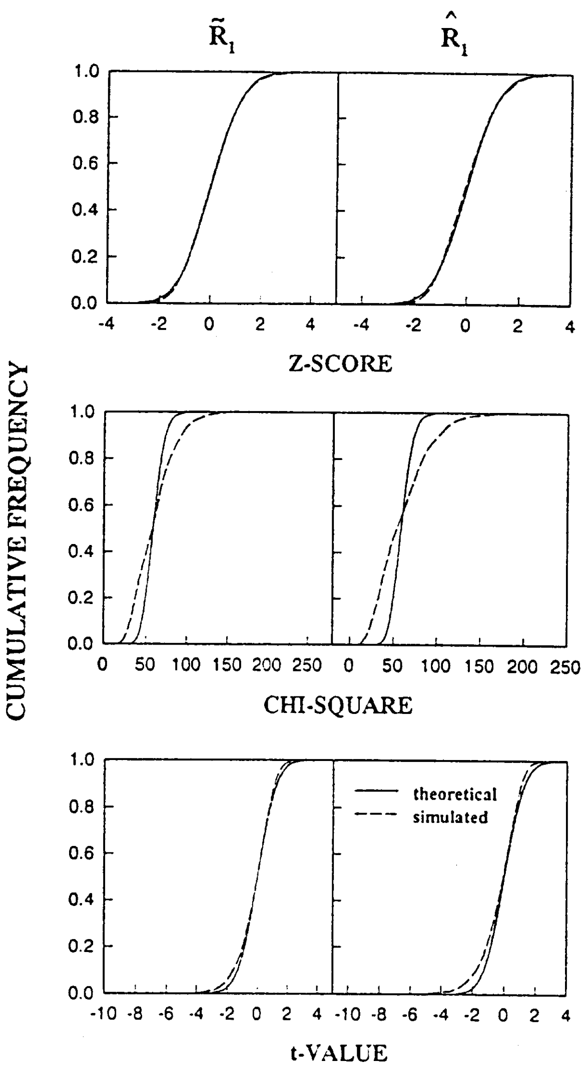


FIGURE 6.—Properties of the distributions of the per-day catch rate estimators $\hat{R}_1$ and $\tilde{R}_1$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerators (z), denominators (χ^2), and t-values for samples of 60 interviews.

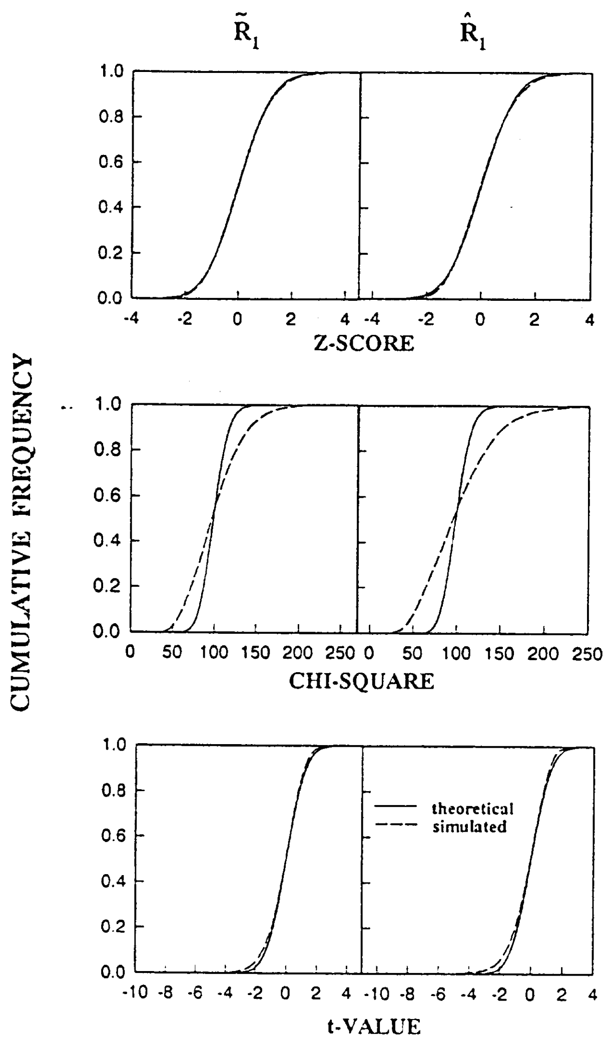


FIGURE 7.—Properties of the distributions of the per-day catch rate estimators $\hat{R}_1$ and $\tilde{R}_1$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerators (z), denominators (χ^2), and t-values for samples of 100 interviews.

tal catch when multiplied by an independent estimate of effort.

There are trade-offs in using $\hat{R}_1$ and $\tilde{R}_1$ to estimate total catch. From the simulations, it appeared that $\tilde{R}_1$ is a better estimator than $\tilde{R}_1$ with respect to symmetry and approach to the targeted α-level. But $\hat{R}_1$ has a smaller SE than $\tilde{R}_1$ and will provide more precise estimates. Under ideal conditions (large sample size and coefficient of variation less than 0.10), both estimators should result in the same estimate of total catch. However, $\hat{R}_1$ is rarely used in an access point survey because total daily catch can usually be estimated by direct expansion, a preferred method that avoids the pitfalls inherent in the use of a ratio estimator.

Skewness of the catch rate distributions seriously affected the confidence interval coverage of the catch rate measures R_1 and R_2, even with sample sizes of 30 and greater—sample sizes at which normality is usually anticipated. To estimate true daily catch rates with approximately 95% confidence, at least 100 anglers must be interviewed. When we simulated at least 100 angler interviews and set a target α-level of 0.05, the confidence bands included the true mean catch rate (R) 95% of the time. At small sample sizes, the realized confidence bands were consistently liberal (we expected a 95% confidence interval, but actually got

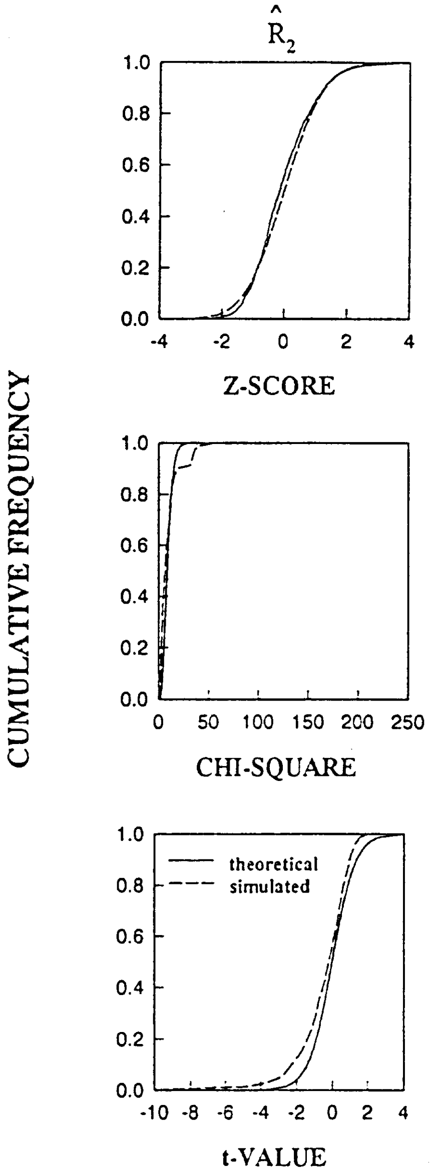


FIGURE 8.—Properties of the distribution of the per-angler catch rate estimator $\hat{R}_2$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerator (z), denominator (χ^2), and t-value for samples of 10 interviews.

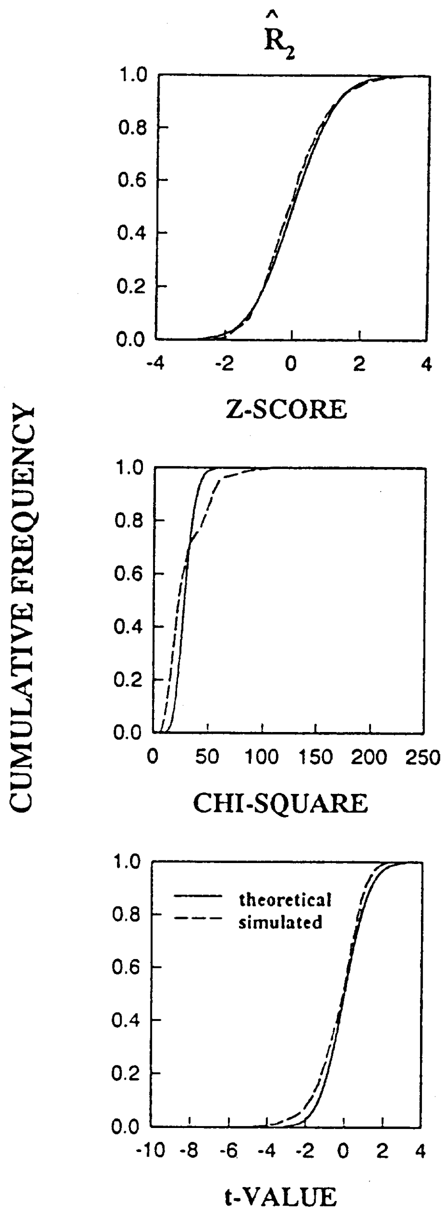


FIGURE 9.—Properties of the distribution of the per-angler catch rate estimator $\hat{R}_2$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerator (z), denominator (χ^2), and t-value for samples of 30 interviews.

a 90% interval). However, we have also shown that the rejection level was not symmetrical and that the upper confidence limit on R was in error much more frequently than the lower limit. This does not affect the width of the confidence bounds, but it will affect their symmetry. In those cases when the confidence band does not include R, the band is more likely to be to the left of R and will underestimate true catch rate—and therefore lead to underestimates of total catch.

In conclusion, we presented three procedures to estimate catch rates: two estimators of the per-day catch rate ($\hat{R}_1$ and $\tilde{R}_1$) and the estimator of the per-angler catch rate ($\hat{R}_2$). When catch rate is obtained from completed-trip interviews of anglers (who are sampled with equal probability), the estimator that is used, $\hat{R}_1$, is the slowest to approach targeted confidence coverage, but it has the smallest SE and the catch rate estimate will be closer to the true catch rate. When catch rate is obtained from uncompleted-trip interviews (i.e., when a fishing trip is interrupted during a roving interview and the probability of sampling that party is propor-

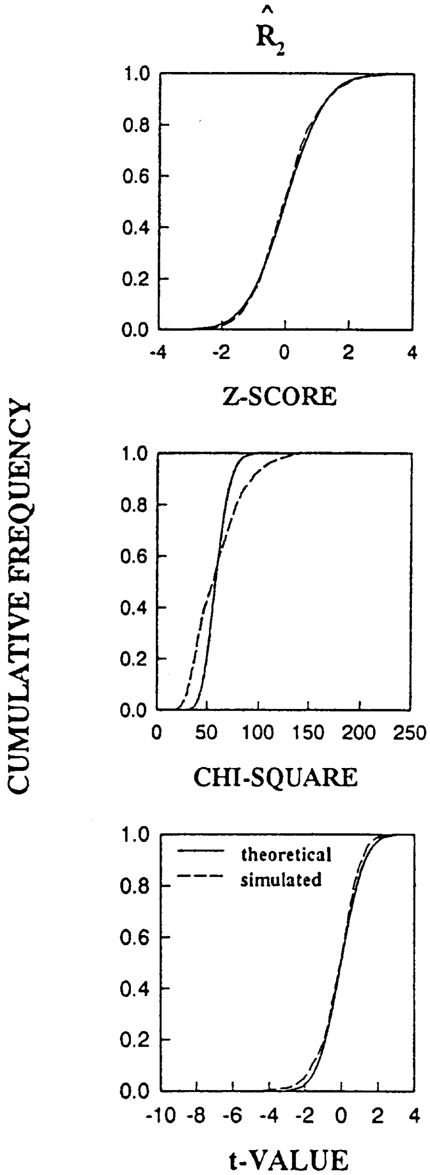


FIGURE 10.—Properties of the distribution of the per-angler catch rate estimator $\hat{R}_2$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerator (z), denominator (χ^2), and t-value for samples of 60 interviews.

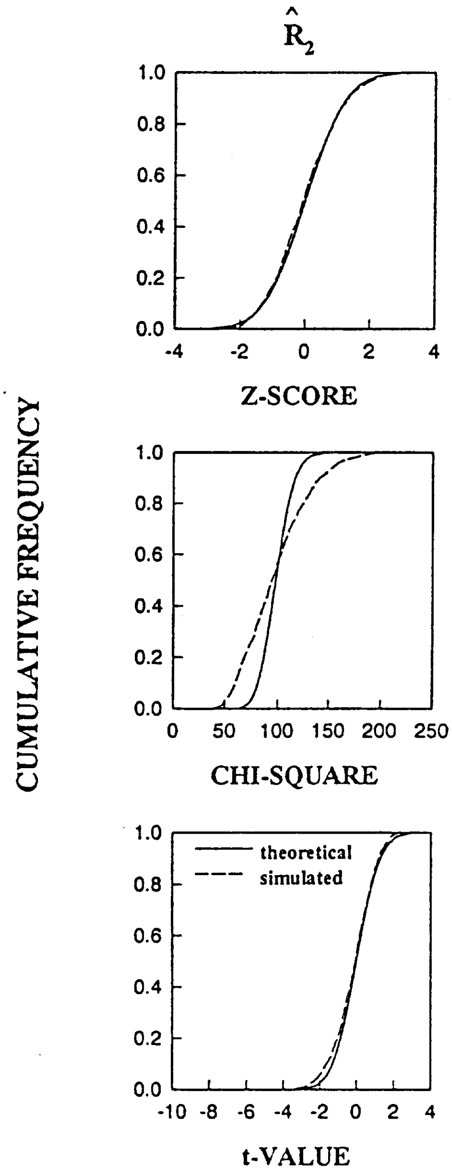


FIGURE 11.—Properties of the distribution of the per-angler catch rate estimator $\hat{R}_2$. Simulated cumulative frequency distributions are compared with the associated theoretical distributions of the numerator (z), denominator (χ^2), and t-value for samples of 100 interviews.

tional to the trip length), the estimator that is used, $\tilde{R}_1$, not only has a larger SE, but its variance can be unstable when very short trips with comparatively large catches are encountered (see equation 5). To date the problems with this estimator have not been resolved in the fisheries literature. The third estimator, $\hat{R}_2$, the statistic that gives the manager information about the characteristics of anglers' success rate, is more robust than the other two with respect to skewness in the catch rate distribution, and it approaches the targeted confidence coverage more quickly than the others.

Acknowledgments

Initial work on this project was funded through a grant from the Chesapeake Bay Stock Assessment Committee and the National Marine Fisheries Service (NA89EA-H-00060) to C. M. Jones. Analysis of the ratio of means estimator with sampling proportional to trip length was funded by a grant from the Virginia Sea Grant Program, (VGMSC-R1MG-91-2) to C. M. Jones. We thank V. Crecco for his review of the initial draft and two anonymous reviewers for their help with the submitted manuscript.

References

Brown, B. E. 1971. Implications for the Oklahoma State Lake Creel Survey to improve creel survey design. Proceedings of the Annual Conference Southeastern Association of Game and Fish Commission 24(1970):577–591.

Cochran, W. G. 1977. Sampling techniques, 3rd edition. Wiley, New York.

Crecco, V. 1989. Recommendations for estimating catch and effort statistics for the marine recreational fishery in Connecticut. Technical Report to the Atlantic States Marine Fisheries Commission, Connecticut Marine Fisheries, Connecticut Department of Environmental Protection, Waterford.

Crone, P. R., and S. P. Malvestuto. 1991. Comparison of five estimators of fishing success from creel survey data on three Alabama reservoirs. American Fisheries Society Symposium 12:61–66.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow. 1953. Sample survey methods and theory, volume 2. Wiley, New York.

Hayne, D. W. 1991. The access point creel survey: procedures and comparison with the roving-clerk creel survey. American Fisheries Society Symposium 12: 123–138.

Jones, C. M., K. Pollock, A. Ehtisham, and W. Hinkle. 1990. Virginia's black drum fishery. Old Dominion University Research Foundation, Technical Report ODURF 90-2, Norfolk, Virginia.

Krebs, C. J. 1989. Ecological methodology. Harper and Row, New York.

Malvestuto, S. P. 1983. Sampling the recreational fishery. Pages 397–419 *in* L. A. Nielsen and D. L. Johnson, editors. Fisheries techniques. American Fisheries Society, Bethesda, Maryland.

Pollock, K. H., C. M. Jones, and T. L. Brown. 1994. Angler survey methods and their applications in fisheries management. American Fisheries Society Special Publication 25.

Robson, D. S. 1991. The roving creel survey. American Fisheries Society Symposium 12:19–24.

Sukhatme, P. V., and B. V. Sukhatme. 1970. Sampling theory of surveys with applications. Iowa State University Press, Ames.

Thompson, S. K. 1992. Sampling. Wiley, New York.

Received May 13, 1992
Accepted April 11, 1995

Appendix 1: Derivation of the Per-Angler Catch Rate Estimator $\hat{R}_2$ for Unequal Probability Sampling

Let Y denote total catch and X denote total effort;

$$Y = \sum_{i=1}^{N} y_i \quad \text{and} \quad X = \sum_{i=1}^{N} x_i.$$

Sampling is with replacement and sampling probability is proportional to trip length; weightings (w) are

$$w_i = \frac{x_i}{\sum_{i=1}^{N} x_i} = \frac{x_i}{X}.$$

If X were known, the estimate of the total catch would be given by

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{w_i},$$

which is an unbiased estimator of Y (Cochran 1977:252). Thus, the expected value of $\hat{Y}$ is Y, or

$$E(\hat{Y}) = Y, \tag{A1.1}$$

where E denotes the mathematical expectation operator. By substitution,

$$E(\hat{Y}) = E\left(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{w_i}\right) = E\left(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{(x_i/X)}\right) = E\left(\frac{1}{n} \sum_{i=1}^{n} X \frac{y_i}{x_i}\right) = X \cdot E\left(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{x_i}\right). \tag{A1.2}$$

The last expression in parentheses is the per-angler catch rate estimator $\hat{R}_2$. From equations (A1.1) and (A1.2), we obtain

$$X \cdot E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}\right) = Y,$$

and it follows that

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}\right) = E(\hat{R}_2) = \frac{Y}{X} = R_1.$$

For the case of unequal probability sampling, we change the name of this estimator from $\hat{R}_2$ to $\tilde{R}_1$ to indicate the parameter it really estimates. Thus,

$$E(\tilde{R}_1) = R_1.$$

Appendix 2: Properties of the Per-Day Catch Rate Parameter R_1

Again, let $Y = \sum_{i=1}^{N} y_i$ be total catch and $X = \sum_{i=1}^{N} x_i$ be total fishing effort. The per-day catch rate is defined as $R_1 = \frac{Y}{X}$. When sampling takes place such that selection probabilities are equal among anglers, the mean catch is $\mu_y = \frac{Y}{N}$, the mean effort is $\mu_x = \frac{X}{N}$, and $R_1 = \frac{\mu_y}{\mu_x}$.

When sampling takes place such that selection probabilities are proportional to trip length, sample weightings (w) are $w_i = x_i/X$; given $\mu_{x,w} = \sum_{i=1}^{N} w_i x_i$, $\mu_{y,w} = \sum_{i=1}^{N} w_i y_i$. Then, the estimator $\hat{R}_1$ gives a consistent but biased estimate of $R_3 = \frac{\mu_{y,w}}{\mu_{x,w}}$.

For R_1, we can show that the variance and covariance terms can be expressed as

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2, \qquad \sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_y)^2, \quad \text{and}$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \equiv \rho_{xy}\sigma_x\sigma_y .$$

The relationship between R_1 and R_3 can be expressed as follows. By definition, coefficients of variation are

$$C_x \equiv \frac{\sigma_x}{\mu_x} \quad \text{and} \quad C_y \equiv \frac{\sigma_y}{\mu_y}; \qquad \text{the regression slope is } \beta \equiv \rho_{xy}\frac{\sigma_y}{\sigma_x}.$$

It can be shown that

$$R_3 = R_1\frac{1 + \rho_{xy}C_xC_y}{1 + C_x^2}.$$

The difference between R_1 and R_3

$$R_1 - R_3 = \frac{C_x^2}{1 + C_x^2}(R_1 - \beta) = \frac{\delta}{\mu_x}\frac{C_x^2}{1 + C_x^2}$$

thus has the same sign and proportional magnitude as the intercept δ of the OLS regression of catch on effort.

Appendix 3: Components of the Simulated t-Value

Transformation of empirical variables to t-distributions.—Traditionally, t is distributed according to a theoretical Student's t-distribution (noted here as t_{th}) with $n - 1$ degrees of freedom, which has the form

$$t_{\text{th}} = \frac{z}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}}, \tag{A3.1}$$

where z is normally distributed with mean 0 and variance 1 and $\chi^2_{(n-1)}$ is independent of z.

Random variates obtained from our simulation models were transformed into t-values and compared with t_{th}, by transforming the numerator into a z-distributed variate and the denominator into a χ^2-distributed variate, as seen in equation (A3.1). These transformations also permit us to examine the numerator and denominator separately. The transformations are obtained by dividing the numerator and denominator by the SD of the numerator:

$$\text{for } \hat{R}_1,\ t_1 = \frac{\dfrac{\bar{y} - R_1\bar{x}}{\sqrt{V(\bar{y} - R_1\bar{x})}}}{\dfrac{\text{SE}(\bar{y} - R_1\bar{x})}{\sqrt{V(\bar{y} - R_1\bar{x})}}}, \quad \text{where } V(\bar{y} - R_1\bar{x}) = \frac{\sum_{i=1}^{N}(y_i - R_1x_i)^2}{nN};$$

$$\text{for } \bar{R}_1,\ t_2 = \frac{\dfrac{\bar{R}_1 - R_1}{\sqrt{V(\bar{R}_1)}}}{\dfrac{\text{SE}(\bar{R}_1)}{\sqrt{V(\bar{R}_1)}}}, \quad \text{where } V(\bar{R}_1) = \frac{\sum_{i=1}^{N} x_i\left(\dfrac{y_i}{x_i} - R_1\right)^2}{n\sum_{i=1}^{N} x_i};$$

$$\text{for } \hat{R}_2,\ t_3 = \frac{\dfrac{(\hat{R}_2 - R_2)}{\sqrt{V(\hat{R}_2)}}}{\dfrac{\text{SE}(\hat{R}_2)}{\sqrt{V(\hat{R}_2)}}}, \quad \text{where } V(\hat{R}_2) = \frac{\sum_{i=1}^{N}\left(\dfrac{y_i}{x_i} - R_2\right)^2}{nN}.$$

The simulated t-values can be compared to t_{th} (with $n - 1$ degrees of freedom) in behavior of the transformed numerator and denominator. The numerator should converge to z for large n (by the central limit theorem), whereas the squared denominator should conform to a χ^2 random variable, $\chi^2_{(n-1)}$. The exact formulations are given below.

Comparison of t_1, t_2, and t_3 with t_{th}.—The numerator of t_1 becomes

$$\frac{\bar{y} - R_1\bar{x}}{\sqrt{V(\bar{y} - R_1\bar{x})}},$$

and the denominator, when squared and multiplied by $n - 1$, is

$$(n-1)\frac{\text{SE}^2(\bar{y} - R_1\bar{x})}{V(\bar{y} - R_1\bar{x})},$$

which is nominally distributed as $\chi^2_{(n-1)}$.

The numerator of t_2 becomes

$$\frac{\bar{R}_1 - R_1}{\sqrt{V(\bar{R}_1)}},$$

and the denominator, when squared and multiplied by $n - 1$, is

$$\frac{(n-1)\mathrm{SE}^2(\tilde{R}_1)}{V(\tilde{R}_1)},$$

which is nominally distributed as χ^2_{n-1}.

The numerator of t_3 becomes

$$\frac{\hat{R}_2 - R_2}{\sqrt{V(\hat{R}_2}} = \frac{\dfrac{\sum_{i=1}^{n} \dfrac{y_i}{x_i}}{n} - R_2}{\sqrt{V(\hat{R}_2)}},$$

and the denominator, when squared and multiplied by $n - 1$, is

$$(n-1)\frac{\mathrm{SE}^2(\hat{R}_2)}{V(\hat{R}_2)},$$

which is nominally distributed as $\chi^2_{(n-1)}$.

TRANSACTIONS OF THE AMERICAN FISHERIES SOCIETY

VOLUME 124, NUMBER 6, NOVEMBER 1995

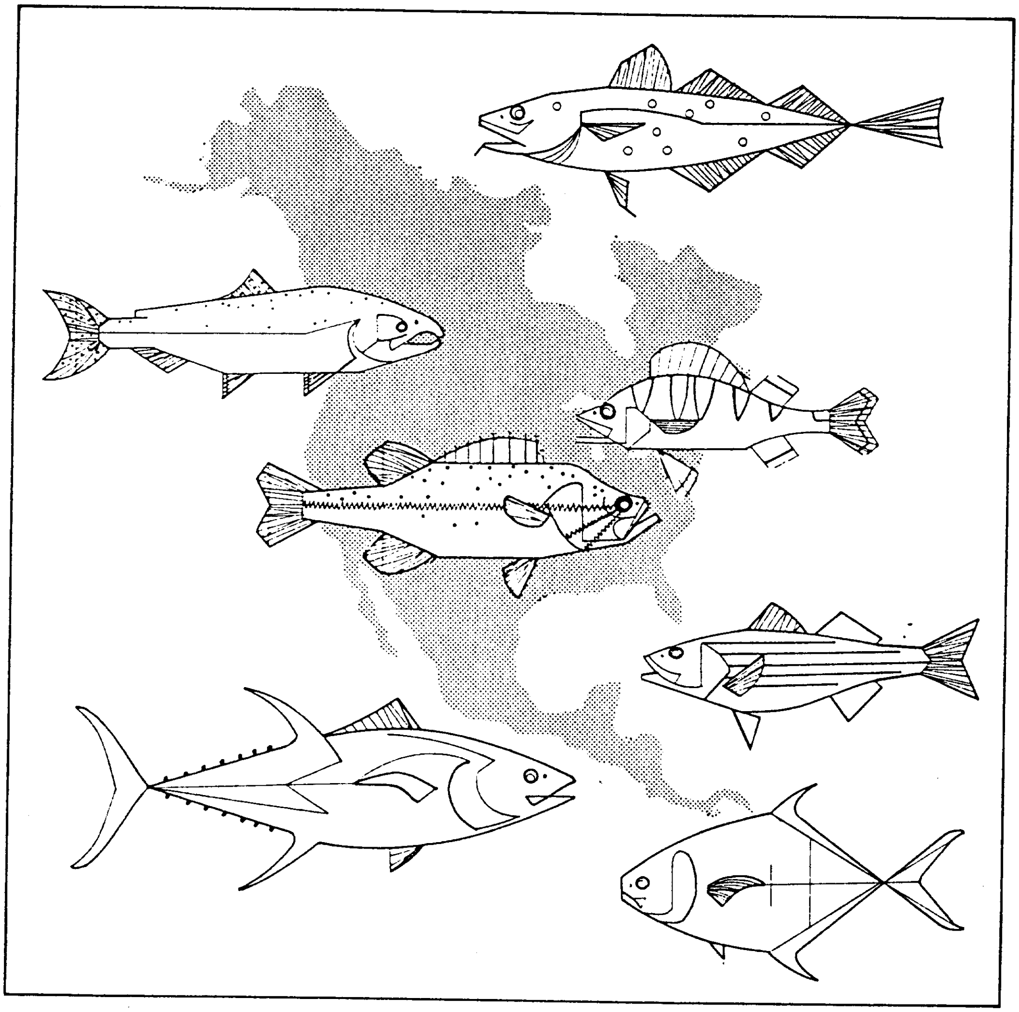
Notes

(Continued on inside back cover)

ISSN 0002-8487 TAFSAI 124(6) 789–973 (1995)

TRANSACTIONS
OF THE AMERICAN FISHERIES SOCIETY

Volume 124 (6) November 1995



QUASQUI-CENTENNIAL YEAR

AMERICAN FISHERIES SOCIETY 1870 - 1995